

보도시점 2023. 8. 7.(월) 조건  
2023. 8. 6.(일) 12:00

배포 2023. 8. 4.(금) 16:00

## “데이터와 인공지능 활용한 신약 개발 경쟁력 도약(JUMP)!”

- 「신약개발 인공지능(AI) 경진대회 ‘도약 인공지능(JUMP AI) 2023’」 개최 -

보건복지부(장관 조규홍)와 과학기술정보통신부(장관 이종호, 이하 ‘과기정통부’)는 8월 7일(월)부터 9월 25(월)까지 「신약개발 인공지능(AI) 경진대회 ‘도약 인공지능(JUMP AI) 2023’」(이하, 신약개발 인공지능(AI) 경진대회)를 개최한다고 밝혔다.

본 경진대회는 한국제약바이오협회, 한국화학연구원, 한국생명공학연구원이 공동 기획하여 주최하였으며, 인공지능(AI) 신약개발에 관심 있는 연구자, 개발자, 학생이 새로운 주제와 데이터를 활용하여 창의적이고 혁신적인 신약개발 인공지능(AI) 모델 개발에 도전토록 함으로써 인공지능(AI) 신약개발 대전환을 가속화 할 경진대회라는 데 의의가 있다.

또한 경진대회에서 ‘화합물 대사안정성 예측모델 개발’에 활용될 데이터는 과기정통부를 중심으로 관계부처가 육성 중인 14대 바이오 소재 클러스터의 합성화합물 분야 한국화합물은행에서 독자 생산한 신뢰도가 높은 고유 연구소재 특성 데이터로서 본 경진대회를 위해 처음으로 개방하여 연구자들에게 인공지능(AI) 신약개발 관련 새로운 고가치 데이터를 활용할 수 있는 기회가 제공된다.

본 경진대회와 관련하여, 일반인을 대상으로 ‘화합물 데이터의 안전한 공유 및 믿고 쓸 수 있는 데이터 활용 등’에 대한 아이디어 공모전도 개최하며, 우수자에게는 별도의 상품도 제공할 예정이다.

본 경진대회는 8월 7일(월)부터 9월 25일(월)까지 신약개발과 인공지능 분야에 관심이 있는 한국 거주 연구자(외국인 포함) 및 재외 한국인은 누구나 AI 신약개발 온라인 교육플랫폼 LAIDD(laidd.org) 및 국가 데이터 스테이션

(K-BDS) (<https://kdbsc.kisti.re.kr>)을 통해 접속 가능하다. 또한, 전산 인프라가 필요할 경우 국가 바이오 데이터 스테이션(K-BDS)을 통해 지원\* 받을 수 있다.

\* 국가바이오데이터스테이션(K-BDS) 데이터 활용지원 포털(<https://kdbsc.kisti.re.kr>)에 회원가입 후, 상담 및 기술지원 요청으로 신청 가능. 자세한 내용은 홈페이지 (dacon.io) 참조.

참가자(팀)는 인간과 쥐의 간 세포에 대한 화합물의 대사안정성을 실험한 학습용 데이터 3,498종을 제공받아 예측모델을 개발하고, 평가용 데이터 490종을 이용하여 예측결과를 제출하게 된다. 9월 말에 참가자(팀) 중 실제 실험값과 가장 가깝게 잘 예측하는 10팀을 선정하고, 10월에 2차 온라인 발표평가를 통해 최종 5팀을 선발할 예정이다.

본 경진대회의 수상자는 10월 23일(월) 발표할 예정이며, 최우수상(장관상) 2점(상금 1,000만 원), 우수상(주최기관장상, 한국제약바이오협회장·한국화학연구원장·한국생명공학연구원장) 3점(상금 300만 원)을 수여한다.

보건복지부 은성호 첨단의료지원관은 “인공지능(AI)을 활용한 신약개발은 우리나라 제약바이오산업 발전의 중대한 기회” 라고 강조하며, “현장밀착형 교육과 경진대회를 통해 양성·발굴된 융합인재는 신약개발 경쟁력 강화와 글로벌 제약기업과의 격차 해소에 핵심적인 역할을 할 것” 이라고 밝혔다.

과기정통부 구혁채 기초원천연구정책관은 “바이오 데이터와 인공지능(AI)의 융합은 연구 및 사업화 과정의 시간과 비용을 크게 절감하고 신산업을 창출하는 디지털바이오의 핵심 요소” 라며, “이번 경진대회를 통해 디지털 바이오가 활성화되고 신약개발 기술 산업에서 글로벌 경쟁력 확보의 발판이 되기를 기대한다” 라고 밝혔다.

자세한 내용은 2023 신약개발 인공지능(AI) 경진대회 플랫폼(<https://url.kr/f3wvaq>)을 통해 확인할 수 있다.

공동 담당 부처	보건복지부 보건의료데이터진흥과	책임자	과 장	심은혜 (044-202-2940)
		담당자	사무관	이은주 (044-202-2947)
공동 담당 부처	과학기술정보통신부 첨단바이오기술과	책임자	과 장	김영혜 (044-202-6180)
		담당자	사무관	장석기 (044-202-6183)



- <붙임> 1. 「신약개발 AI 경진대회」 개요  
 2. 경진대회 활용 데이터  
 3. 경진대회 포스터

**□ 대회개요**

- (대회명) 2023 신약개발 인공지능(AI) 경진대회, “도약 인공지능(JUMP AI)”
- (주최) 한국제약바이오협회, 한국화학연구원, 한국생명공학연구원
- (주관) 인공지능(AI)신약개발지원센터, 한국화합물은행, 국가생명연구자원정보센터
- (후원) 보건복지부, 과학기술정보통신부
- (신청기간) 2023. 8. 7.(월) ~ 9. 25.(월)
- (참가대상) 한국 거주자(외국인 포함) 및 재외 한국인

**□ 대회일정**

- (경진대회 개최) '23.8.7.(월) 10:00 ~ 9.25.(월) 10:00
- (코드 및 발표자료 제출 마감) '23.10.06.(금) 23:59
- (코드 평가) '23.10.9.(월) ~ 10.20.(금)
- (온라인 발표평가) '23.10.13.(금)(온라인 ZOOM 활용)
- (결과 발표) '23.10.23.(월) 10:00
- (포상 수여) '23년 AI Pharma korea 컨퍼런스에서 시상('23.11월 초 예정)

## □ 심사절차

### ○ 정량평가점수와 정성평가점수(발표평가)를 6:4로 반영

#### - (정량평가)

- ▶인간의 간 대사효소(HLM: human liver microsome) 및 실험용 쥐의 간 대사효소(MLM: mouse liver microsome)에 대한 학습용 데이터(training set, 3,498종, 화학구조 정보 및 대사안정성 측정값 제공, 논문 등 검증된 외부 데이터 활용 가능)를 제공하고 참가자들이 이를 바탕으로 예측모델 개발
- ▶개발한 예측 모델로 경진용 데이터(Test set, 490종, 화학구조 정보만 제공)를 사용하여 예측한 값을 행사 사이트에 제출
- ▶제출된 예측값과 한국화학물은행의 실측값과 비교하여 평균제곱근 편차(RMSE: root mean square) 값으로 예측 정확도 평가

#### - (정성평가) 모델 수립의 정성적 우수성 평가(심사위원 3 ~5인)

항목	심사기준	점수
리더보드 private 점수	모델 예측 성능	60
모델링 전략	도메인을 고려한 모델 활용	10
데이터 전처리 및 validation 데이터 구축 전략	데이터 특성 고려한 전략	10
실험관리	다양한 실험 시도와 논리성	10
적용성(모델의 범용성)	모델의 다른 분야 적용 가능성	10

## □ 시상 내역

- 보건복지부 장관상, 과학기술정보통신부 장관상, 주최 기관장상 및 협회장상으로 상금 총 2,900만 원

순위	표창	인원	상금
최우수상	보건복지부 장관상/ 과학기술정보통신부 장관상	2인(팀)	각 1,000만 원
우수상	한국제약바이오협회장상/ 한국화학연구원장상 / 한국생명공학연구원장상	3인(팀)	각 300만 원
합계		5인(팀)	총 2,900만 원

## □ 대사 안정성 데이터

- (정의) 인간 및 마우스의 간 대사효소와 화합물을 30분 동안 반응시킨 후, 대사되지 않고 남아있는 화합물의 양을(%) 질량 분석(LC-MS/MS) 방법으로 측정함으로써 화합물의 간 대사효소에 대한 안정성을 평가한 데이터
- (경진대회 활용 데이터) 한국화합물은행 대표 라이브러리\* 중 화합물 약 3,988종에 대하여 수행한\*\* 인간/실험용 쥐의 간 마이크로솜 대사안정성 평가(human/mouse liver microsomal stability assay) 데이터
  - 학습용 데이터(Training set, 3,498종 화합물), 경진용 데이터(test set, 490종 화합물)
    - \* 한국화합물은행 보유 전체 화합물 구조의 특성을 대표할 수 있는 라이브러리
    - \*\* 화학(연) 정보융합신약연구센터에서 분석 수행

%remaining (30 min, MLM)	총합계 (%)	%remaining (30 min, HLM)	총합계 (%)
0-10	1514 (40.0)	0-10	820 (20.6)
10-20	286 (7.2)	10-20	251 (6.3)
20-30	255 (6.4)	20-30	240 (6.0)
30-40	214 (5.4)	30-40	210 (5.3)
40-50	237 (5.9)	40-50	243 (6.1)
50-60	225 (5.6)	50-60	258 (6.5)
60-70	223 (5.6)	60-70	307 (7.7)
70-80	276 (6.9)	70-80	365 (9.2)
80-90	276 (6.9)	80-90	384 (9.6)
90-100	482 (12.1)	90-100	910 (22.8)
총합계	3988	총합계	3988

Model	MLM			HLM		
	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>
Linear Regression	24.04	29.19	0.33	25.01	29.40	0.31
Decision Tree (Vanilla)	30.66	40.26	-0.26	34.56	44.24	-0.55
Random Forest (Vanilla)	25.36	29.74	0.30	25.53	29.89	0.29
GAT	21.02	26.29	0.45	21.95	26.55	0.43
GCN	22.45	28.20	0.38	23.40	28.55	0.37
MPNN	20.57	25.74	0.48	21.83	27.20	0.43
Attentive FP	20.74	25.33	0.50	21.97	26.36	0.45

<약 4천 종 화합물에 대한 대사안정성 분석 결과>

## □ 데이터의 가치

- (약물의 대사안정성) 인체 내에서 약물 노출에 영향을 주는 주요 인자로 약효와 독성을 모두 변화시킬 수 있어, 대사안정성 평가 데이터의 확보는 신약개발의 후보 물질 최적화 및 선별 과정에 필수
  - 논문 등에 발표되지 않은 소재와 데이터로써 가치가 매우 높음
- (화합물의 구조 다양성) 데이터 증강기술로 보완이 가능하며, 화합물 구조의 다양성으로 예측 모델의 과적합(overfitting) 가능성 경감.
- (데이터의 균질성 및 신뢰도) 동일한 실험실에서 일정 조건으로 데이터 생산
- (기대효과) 데이터를 활용한 대사 안정성의 예측 모델 개발은 신약개발 효율성 제고에 기여

JUMP AI 2023

# 제1회 신약개발 AI 경진대회 대사안정성 예측

Quantum Jump by AI for  
Drug Discovery & Development

## I 주제

화합물 대사안정성 예측 모델 개발  
한국화합물은행의 고유 대사안정성데이터 제공

## I 목적

국내 신약 소재 데이터의 공유·활용을 활성화하고,  
창의적이고 혁신적인 AI융합인재를 발굴

## I 대회 일정

2023.8.7(월) ~ 2023.10.23(월)

## I 참가대상

한국 거주 연구자(외국인포함) 및 재외 한국인

## I 문의

대회 진행: dacon@dacon.ai

## I 상세 일정

대회 시작	08월 07일 (월)
대회 종료	09월 25일 (월)
코드 및 발표자료 제출 마감	10월 06일 (금)
코드 평가	10월 09일 (월) ~ 10월 20일 (금)
온라인 발표평가	10월 13일 (금)
결과 발표	10월 23일 (월)
시상식	11월 초 (수상자 별도안내)

## I 대회 상금

최우수상	보건복지부장관상 과학기술정보통신부장관상	각 1,000만원
우수상	한국제약바이오협회장상 한국화학연구원장상 한국생명공학연구원장상	각 300만원

\*제세공과금은 개인 부담

\*세부 상금은 대회 운영상황에 따라 변동될 수 있습니다.

대회  
참여



주최·주관 KPBMA 한국제약바이오협회 KRICT 한국화학연구원 KRISS 한국생명공학연구원

후원 보건복지부 과학기술정보통신부